

# Multiscale Granger Causality

L. Faes and G. Nollo

*Bruno Kessler Foundation, Trento, Italy and  
BIOtech, Dept. of Industrial Engineering, University of Trento, Italy*

S. Stramaglia

*Dipartimento di Fisica, Università degli Studi Aldo Moro, Bari, Italy and  
INFN, Sezione di Bari, Italy*

D. Marinazzo

*Data Analysis Department, Ghent University, Ghent, Belgium  
(Dated: March 27, 2017)*

In the study of complex physical and biological systems represented by multivariate stochastic processes, an issue of great relevance is the description of the system dynamics spanning multiple temporal scales. While methods to assess the dynamic complexity of individual processes at different time scales are well-established, the multiscale evaluation of directed interactions between processes is complicated by theoretical and practical issues such as filtering and downsampling. Here we extend the very popular measure of Granger causality (GC), a prominent tool for assessing directed lagged interactions between joint processes, to quantify information transfer across multiple time scales. We show that the multiscale processing of a vector autoregressive (AR) process introduces a moving average (MA) component, and describe how to represent the resulting ARMA process using state space (SS) models and to combine the SS model parameters for computing exact GC values at arbitrarily large time scales. We exploit the theoretical formulation to identify peculiar features of multiscale GC in basic AR processes, and demonstrate with numerical simulations the much larger estimation accuracy of the SS approach compared with pure AR modeling of filtered and downsampled data. The improved computational reliability is exploited to disclose meaningful multiscale patterns of information transfer between global temperature and carbon dioxide concentration time series, both in paleoclimate and in recent years.

PACS numbers: 02.50.Ey, 87.10.Mn, 92.70.Gt

Granger causality (GC) is a powerful tool for assessing directional interactions from time series data according to the notion of time lagged influence first proposed by Wiener [1] and then formalized by Granger [2] and Geweke [3, 4] in the framework of vector autoregressive (AR) modeling of stochastic processes. Since its formulation, GC has gained increasing popularity and is nowadays ubiquitously employed in several scientific fields ranging from econometrics to social and climate sciences, neuroscience and physiology [5–8]. The great success of this measure comes from its conceptual simplicity, data-driven nature, and relative ease of implementation. An additional appealing property of GC is the principled interpretation of its generalized probabilistic formulation, which is closely related to the information-theoretic concept of transfer entropy [9].

Since many dynamic processes in physics, biology and other fields display peculiar features deployed across multiple temporal scales [10–14], it is likely that the information flowing between subsystems also exhibit a multiscale structure. The most common approach to investigate multiscale dynamics is to resample at different temporal scales the originally measured realization of an observed stochastic process, yielding a collection of rescaled time series from which various dynamical measures can be calculated [15]. While this approach has been followed

mostly to quantify the multiscale behavior of the individual dynamics of scalar processes [11–14], attempts have been made to extend it to the computation of the information transferred between processes at multiple scales [16, 17]. However, the multiscale evaluation of lagged influence measures such as the GC is severely complicated by theoretical and practical issues [18, 19]. These issues arise from the rescaling procedure, which essentially consists in a filtering step eliminating the fast temporal scales (classically performed by averaging [15]) followed by a downsampling step that coarse-grains the time series around the selected scale. The filtering step leaves theoretically unchanged the GC values, but degrades severely their estimation affecting reliability, stability and data demand [20]. The downsampling step is even more problematic, as it alters GC values in a way that was unknown until very recently [18, 19] and impacts consistently detectability and accuracy of GC estimates.

Thus, a reliable assessment of multiscale GC must provide a theoretical, rather than empirical, approach to derive this measure after rescaling the observed processes at any given temporal scale. To this end, here we propose an analytical frame for the computation of GC for linear multivariate stochastic processes subjected to averaging and downsampling. The framework exploits the theory of state space (SS) models [21], and builds on re-

cent theoretical results [18, 19] to yield exact GC values for coupled processes observed at different time scales. In this study, it is illustrated for examples of coupled AR processes and applied to assess, for the first time spanning a very large range of temporal scales, the patterns of information transfer between anthropogenic emissions and global temperatures.

To lay the groundwork for multiscale GC computation and introduce notations, we start resuming the calculation of GC for general multivariate processes [2, 4] and for state space (SS) processes [18]. Let us consider a discrete-time, stationary vector stochastic process composed of  $M$  real-valued zero-mean scalar processes,  $Y_n = [y_{1,n} \cdots y_{M,n}]^T$ ,  $-\infty < n < \infty$ , and assume the process  $y_j$  as the *target* and the process  $y_i$  as the *driver* (the remaining  $M - 2$  processes form the vector  $Y_k$ , where  $k = \{1, \dots, M\} \setminus \{i, j\}$ ). Then, denoting the present and the past of vector and scalar variables respectively as  $Y_n$ ,  $y_n$ , and  $Y_n^- = [Y_{n-1}^T Y_{n-2}^T \cdots]$ ,  $y_n^- = [y_{n-1} y_{n-2} \cdots]$ , GC from  $y_i$  to  $y_j$  (conditional on  $Y_k$ ) quantifies the extent to which  $y_{i,n}$  improves the prediction of  $y_{j,n}$  above and beyond the extent to which  $y_{j,n}$  is predicted by  $y_{j,n}^-$  and  $Y_{k,n}^-$ . This definition is assessed in the time domain performing a regression of the present of the target on the past of all processes, yielding the prediction error  $e_{j|ijk,n} = y_{j,n} - \mathbb{E}[y_{j,n} | Y_n^-]$ , and on the past of all processes except the driver, yielding the prediction error  $e_{j|jk,n} = y_{j,n} - \mathbb{E}[y_{j,n} | y_{j,n}^-, Y_{k,n}^-]$  ( $\mathbb{E}$  is the expectation operator). The prediction error variances resulting from these "full" and "restricted" regressions,  $\lambda_{j|ijk} = \mathbb{E}[e_{j|ijk,n}^2]$  and  $\lambda_{j|jk} = \mathbb{E}[e_{j|jk,n}^2]$  are then combined to yield GC from  $y_i$  to  $y_j$  as [4]

$$F_{i \rightarrow j} = \ln \frac{\lambda_{j|jk}}{\lambda_{j|ijk}}. \quad (1)$$

The measure (1) is the log-likelihood ratio for the two linear regressions associated with the projections  $\mathbb{E}[y_{j,n} | Y_n^-]$  and  $\mathbb{E}[y_{j,n} | y_{j,n}^-, Y_{k,n}^-]$  [18], and has interpretation as the rate of "information transfer" from driver to target [9]. It can be calculated efficiently from the SS representation of the observed process  $Y$ , which is defined as [22]

$$X_{n+1} = \mathbf{A}X_n + W_n \quad (2a)$$

$$Y_n = \mathbf{C}X_n + V_n \quad (2b)$$

where  $X$  is the state (unobserved) process, and  $W$  and  $V$  are zero-mean white noise processes with covariances  $\Xi \equiv \mathbb{E}[W_n W_n^T]$  and  $\Psi \equiv \mathbb{E}[V_n V_n^T]$ , and cross-covariance  $\Theta \equiv \mathbb{E}[W_n V_n^T]$ . The SS process has an equivalent representation, referred to as "innovations form" SS (ISS), evidencing the *innovations*  $E_n = Y_n - \mathbb{E}[Y_n | Y_n^-]$  with covariance  $\Phi \equiv \mathbb{E}[E_n E_n^T]$ , and characterized by the state process  $Z_n = \mathbb{E}[X_n | Y_n^-]$ :

$$Z_{n+1} = \mathbf{A}Z_n + \mathbf{K}E_n \quad (3a)$$

$$Y_n = \mathbf{C}Z_n + E_n. \quad (3b)$$

The SS and ISS representations share the state and observation matrices  $\mathbf{A}$  and  $\mathbf{C}$ , and differ in the noise matrices  $(\Xi, \Psi, \Theta)$  and  $(\mathbf{K}, \Phi)$ . To find the ISS parameters  $(\mathbf{A}, \mathbf{C}, \mathbf{K}, \Phi)$  from the SS parameters  $(\mathbf{A}, \mathbf{C}, \Xi, \Psi, \Theta)$  it is necessary to solve a so-called discrete algebraic Ricatti equation (*DARE*), formulated in terms of the state error variance matrix  $\mathbf{P}$ :

$$\mathbf{P} = \mathbf{A}\mathbf{P}\mathbf{A}^T + \Xi - (\mathbf{A}\mathbf{P}\mathbf{C}^T + \Theta)(\mathbf{C}\mathbf{P}\mathbf{C}^T + \Psi)^{-1}(\mathbf{C}\mathbf{P}\mathbf{A}^T + \Theta^T), \quad (4)$$

from which  $\mathbf{K}$  and  $\Phi$  are obtained as

$$\begin{aligned} \Phi &= \mathbf{C}\mathbf{P}\mathbf{C}^T + \Psi \\ \mathbf{K} &= (\mathbf{A}\mathbf{P}\mathbf{C}^T + \Theta)\Phi^{-1}. \end{aligned} \quad (5)$$

Then, GC can be computed from the ISS parameters as follows [18]. The error variance of the full regression is simply the  $j$ -th diagonal element of the innovation covariance,  $\lambda_{j|ijk} = \Phi(j, j)$ . The error of the restricted regression is obtained by forming a *submodel* that excludes the driver process, i.e. a state space model with state equation (3a) and observation equation

$$Y_n^{(jk)} = \mathbf{C}^{(jk)}Z_n + E_n^{(jk)} \quad (6)$$

where the superscript  $(a)$  denotes selection of the rows with indices  $a$  of a matrix. The submodel (3a, 6) is an SS model with parameters  $(\mathbf{A}, \mathbf{C}^{(jk)}, \mathbf{K}\Phi\mathbf{K}^T, \Phi(jk, jk), \mathbf{K}\Phi(:, jk))$ , which can be converted to an ISS model with innovation covariance  $\Phi^R$  solving the *DARE* (4,5), so that the restricted error variance becomes  $\lambda_{j|jk} = \Phi^R(j, j)$ . This shows that GC can be computed numerically from the ISS parameters  $(\mathbf{A}, \mathbf{C}, \mathbf{K}, \Phi)$  of an observed process  $Y$ .

Now we move to develop our framework for multiscale computation of GC. Here we consider the most common operationalization of GC, i.e. that grounded on the AR representation of multivariate processes [2, 7–9]:

$$Y_n = \sum_{k=1}^p \mathbf{A}_k Y_{n-k} + U_n, \quad (7)$$

where  $p$  is the model order,  $\mathbf{A}_k$  are  $M \times M$  matrices of coefficients, and  $U_n = [u_{1,n} \cdots u_{M,n}]^T$  is a vector of  $M$  zero mean Gaussian innovation processes with covariance matrix  $\Sigma \equiv \mathbb{E}[U_n U_n^T]$ . To study the observed process  $Y$  at the temporal scale identified by the scale factor  $\tau$ , we apply the following transformation to each constituent process  $y_m$ ,  $m = 1, \dots, M$ :

$$\bar{y}_{m,n} = \sum_{l=0}^q b_l y_{m,n\tau-l}. \quad (8)$$

This rescaling operation corresponds to transform the original process  $Y$  through a two step procedure that consists of the following *filtering* and *downsampling* steps,

yielding respectively the processes  $\tilde{Y}$  and  $\bar{Y}$ :

$$\tilde{Y}_n = \sum_{l=0}^q b_l Y_{n-l}, \quad (9a)$$

$$\bar{Y}_n = \tilde{Y}_{n\tau}, n = 1, \dots, N/\tau \quad (9b)$$

The change of scale in (8) generalizes the averaging procedure originally proposed in [15], which sets  $q = \tau - 1$  and  $b_l = 1/\tau$ . Here, to improve elimination of the fast temporal scales [23], we identify the  $b_l$  as the coefficients of a linear lowpass filter with cutoff frequency set at  $f_\tau = 1/2\tau$  to avoid aliasing in the subsequent downsampling step [24]. Substituting (7) in (9a), the filtering step leads to the process representation:

$$\tilde{Y}_n = \sum_{k=1}^p \mathbf{A}_k \tilde{Y}_{n-k} + \sum_{l=0}^q \mathbf{B}_l U_{n-l} \quad (10)$$

where  $\mathbf{B}_l = b_l \mathbf{I}_M$  ( $\mathbf{I}_M$  is the  $M \times M$  identity matrix). Hence, the change of scale introduces a moving average (MA) component of order  $q$  in the original  $\text{AR}(p)$  process, transforming it into an  $\text{ARMA}(p, q)$  process. Then, exploiting the close relation between ARMA and SS models [21], the process (10) is turned into an ISS model by defining the state process  $\tilde{Z}_n = [Y_{n-1}^T \dots Y_{n-p}^T U_{n-1}^T \dots U_{n-q}^T]^T$  that, together with  $\tilde{Y}_n$ , obeys the state equations (3) with parameters  $(\tilde{\mathbf{A}}, \tilde{\mathbf{C}}, \tilde{\mathbf{K}}, \tilde{\Phi})$ , where

$$\tilde{\mathbf{C}} = [\mathbf{A}_1 \dots \mathbf{A}_p \mathbf{B}_1 \dots \mathbf{B}_q],$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} \tilde{\mathbf{C}} & & \\ \mathbf{I}_{M(p-1)} & \mathbf{0}_{M(p-1) \times M(q+1)} & \\ \mathbf{0}_{M \times M(p+q)} & & \\ \mathbf{0}_{M(q-1) \times Mp} & \mathbf{I}_{M(q-1)} & \mathbf{0}_{M(q-1) \times M} \end{bmatrix},$$

$$\tilde{\mathbf{K}} = [\mathbf{I}_M \mathbf{0}_{M \times M(p-1)} \mathbf{B}_0^{-T} \mathbf{0}_{M \times M(q-1)}]^T,$$

and where  $\tilde{\Phi} = \mathbf{B}_0 \Sigma \mathbf{B}_0^T$  is the covariance of the innovations  $\tilde{E}_n = \mathbf{B}_0 U_n$ . Moreover, the downsampled process  $\bar{Y}_n$  can be put in ISS form directly from the ISS formulation of the filtered process  $\tilde{Y}_n$ : exploiting a recent result (theorem III in [19]), we find that  $\bar{Y}_n = \tilde{Y}_{n\tau}$  has an ISS representation with state process  $\bar{Z}_n = \tilde{Z}_{n\tau}$ , innovation process  $\bar{E}_n = \tilde{E}_{n\tau}$ , and parameters  $(\bar{\mathbf{A}}, \bar{\mathbf{C}}, \bar{\mathbf{K}}, \bar{\Phi})$ , where  $\bar{\mathbf{A}} = \tilde{\mathbf{A}}^\tau$ ,  $\bar{\mathbf{C}} = \tilde{\mathbf{C}}$ , and where  $\bar{\mathbf{K}}$  and  $\bar{\Phi}$  are obtained solving the *DARE* (4,5) for the SS model  $(\bar{\mathbf{A}}, \bar{\mathbf{C}}, \bar{\Xi}_\tau, \bar{\Phi}, \bar{\Theta}_\tau)$  with

$$\begin{aligned} \bar{\Theta}_\tau &= \tilde{\mathbf{A}}^{\tau-1} \tilde{\mathbf{K}} \tilde{\Phi} \\ \bar{\Xi}_\tau &= \tilde{\mathbf{A}} \bar{\Xi}_{\tau-1} \tilde{\mathbf{A}}^T + \tilde{\mathbf{K}} \tilde{\Phi} \tilde{\mathbf{K}}^T, \tau \geq 2 \\ \bar{\Xi}_1 &= \tilde{\mathbf{K}} \tilde{\Phi} \tilde{\mathbf{K}}^T, \tau = 1. \end{aligned} \quad (11)$$

The overall procedure for multiscale analysis is depicted in Fig. 1: filtering with cutoff  $f_\tau$  the  $\text{AR}(p)$  process  $Y$  yields an  $\text{ARMA}(p, q)$  process, which is equivalent

to an ISS process; the subsequent downsampling yields a different SS process, which in turn can be converted to the ISS form solving the *DARE*. Thus, both filtered and downsampled processes are described by ISS models, whose parameters can be used to compute GC by forming a submodel in which the target is observed without considering the driver process (eq. (6)) and solving the *DARE* for this submodel. This procedure allows analytical computation of GC measures for multiscale (filtered and downsampled) processes, which is illustrated in the following for simulated and real time series.

Theoretical analysis and simulations are performed for the bivariate AR process with equations:

$$y_{1,n} = a_1 y_{1,n-b_1} + c_1 y_{2,n-d_1} + u_{1,n} \quad (12a)$$

$$y_{2,n} = a_2 y_{2,n-b_2} + c_2 y_{1,n-d_2} + u_{2,n} \quad (12b)$$

with iid noise processes  $u_{1,n}, u_{2,n} \sim \mathcal{N}(0, 1)$ . The parameters in (12) are set to generate autonomous dynamics with strength  $a_i$  and lag  $b_i$  for each scalar process  $y_i$ , and causal interactions with strength  $c_i$  and lag  $d_i$  from  $y_j$  to  $y_i$  ( $i, j = 1, 2$ ). We consider two parameter configurations: unidirectional interaction at lag 2 from  $y_1$  to  $y_2$ , obtained setting  $c_1 = 0$  and  $c_2 = 0.5, d_2 = 2$ , where also autonomous dynamics are generated for  $y_1$  but not for  $y_2$  ( $a_1 = 0.5, b_1 = 1, a_2 = 0$ ); bidirectional interactions with different lags and strengths ( $c_1 = 0.75, d_1 = 2; c_2 = 0.5, d_2 = 7$ ) in the presence of autonomous dynamics for both processes ( $a_1 = a_2 = 0.5, b_1 = b_2 = 1$ ). First, we study the exact values of multiscale GC obtained from the true AR parameters. The theoretical trends depicted in Figs. 2 and 3 (black solid lines) document from the perspective of SS modeling the invariance of GC under filtering, already proven in [20]. The behavior of the information transfer across multiple temporal scales is thus shaped by the downsampling step, revealing the tendency of GC to peak at scales corresponding with the lag of the imposed causal interactions: maximal information transfer is found at  $\tau = 2$  for  $F_{1 \rightarrow 2}$  in the unidirectional scheme (Fig. 3a,b), and at  $\tau = 7$  for  $F_{1 \rightarrow 2}$

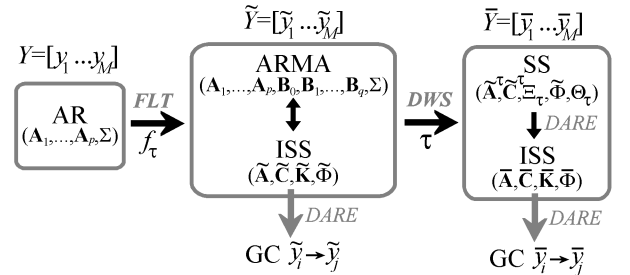


FIG. 1. Schematic representation of a linear multivariate AR process (left) and of its multiscale representation obtained through filtering (FLT) and downsampling (DWS) steps. GC can be computed for any scale factor  $\tau \geq 1$  from the parameters of the ISS representation of the process ( $\tau = 1$  yields GC for the non-rescaled processes).

and  $\tau = 2$  for  $F_{2 \rightarrow 1}$  in the bidirectional scheme (Fig. 3c). The behavior is general, in the sense that it was observed also for different parameter configurations.

Next, we test reliability of multiscale GC estimates obtained from finite length realizations of (12) (a) following a naïve approach whereby GC is computed performing full and restricted regressions on the filtered and downsampled time series, and (b) performing AR identification on the original time series and then applying the new proposed framework to the estimated AR parameters [25]. Application of the two approaches to 100 process realizations of 500 points is depicted in Figs. 2a, 3a and in Figs. 2b, 3b respectively, and indicates the need of state-space analysis: while the computation of GC after filtering and downsampling returns strongly biased and highly variable estimates, the new framework yields accurate detection of GC across multiple scales.

As a practical application, we consider the multiscale analysis of GC between carbon dioxide concentration ( $CO_2$ ) and global temperature ( $GT$ ). It is widely considered that the raise of  $CO_2$  is a main cause of global warming [26], although the validity of such causal relation is still under debate. The problem of understanding the causes of climate change is usually tackled by numerical experiments using Global Climate Models [27] which aim at catching the complexity of climate dynamics. However, data-driven approaches, as GC, are also fruitful in assessing cause-effect relationships between temperature and external forcings. In [28] it has been shown that  $CO_2$  Granger causes temperature, based on data from 1860 to 2008, partly from ice cores, and analyzing second differences of both  $CO_2$  and  $GT$ . Similar conclusions were found in [29], using GC, in [30] by estimating the time rate of information flowing from one time series to the other, and in [6] using a physical approach. Here,

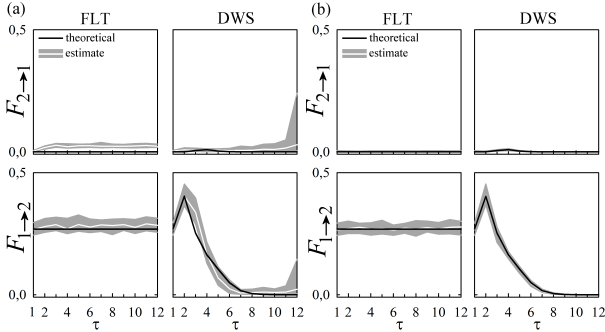


FIG. 2. Multiscale GC analysis of the AR process (12) configured to unidirectional coupling from  $y_1$  to  $y_2$ . Plots depict the theoretical values (black lines) and distribution of estimates (median: white lines; interquartile range: grey areas) of GC computed as a function of the time scale after filtering (FLT) and downsampling (DWS). Estimates are obtained using the naïve AR approach (a) and the proposed framework implemented using a lowpass FIR filter of order  $q = 6$  (b).

we first analyze the global land-ocean temperature index [31] and  $CO_2$  concentration [32] measured at monthly resolution from March 1958 to February 2017 (708 data points) and de-trended applying an L1 norm filter to full-fill stationarity. As depicted in Fig. 4a, the GC along the direction  $GC\ CO_2 \rightarrow GT$  appears to be predominant over the opposite causal influence; the result is evident only spanning temporal scales of several years.

Next, we dramatically change the time scales and turn to consider paleoclimatological data, so as to analyze the GC between  $GT$  and  $CO_2$  concentration on the Vostok Ice Core data from 400,000 to 6,000 years ago, extended by the EPICA Dome C data which go back to 800,000 years ago [33]. In [34], empirical evidences for the existence of Granger causal influences along both directions  $CO_2 \rightarrow GT$  and  $GT \rightarrow CO_2$  have been found after correcting for deterministic trends on the same data. Here, applying our framework for multiscale analysis to the data resampled to a uniform time spacing of 729.77 years (time series length = 1095 points), we obtained the GC curves reported in Fig. 4b. Our results indicate that, at paleolithic time scales, the GC  $GT \rightarrow CO_2$  is largely predominant and peaks at  $\sim 1000$  and  $\sim 10000$  years. These results may be related to the lags between Antarctic deglacial warming and  $CO_2$  increase reported in [35], and also confirm the good evidences reported on the fact that higher global temperatures do promote a rise of greenhouse gas levels [36].

Summarizing, our bivariate analysis of carbon dioxide and temperature changes shows that the two variables are interdependent and interacting at multiple time scales, with a predominance of  $GT \rightarrow CO_2$  effects at paleolithic time scales, and of  $CO_2 \rightarrow GT$  influences at the time scales of modern climate. We remark that our analysis, for the first time to our knowledge, has revealed nontrivial GC  $GT \rightarrow CO_2$  also in modern climate, thus implying a positive feedback which will increase the effect of anthropogenic emissions on global temperatures.

In conclusion, our study opens the way to the theoretical analysis of the information flowing between linear

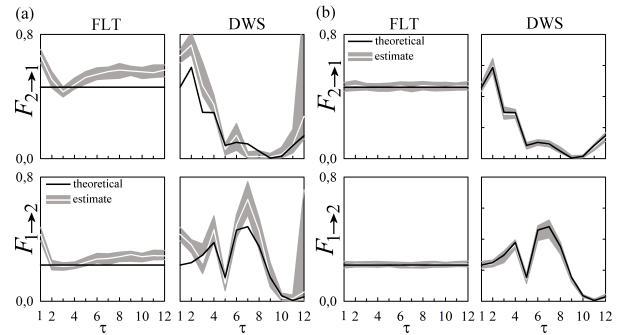


FIG. 3. Multiscale GC analysis of the AR process (12) configured to bidirectional coupling between  $y_1$  and  $y_2$ . Plots and symbols are as in Fig. 2.

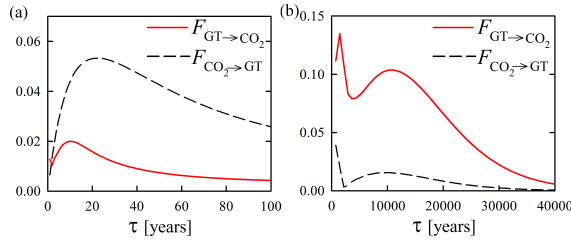


FIG. 4. Multiscale GC analysis of global temperature ( $GT$ ) and  $CO_2$  concentration performed for modern climate (a) and paleoclimate data (b). Computations are performed using the proposed framework implemented with an order-6 FIR lowpass filter. The initial AR model order, set by the BIC criterion, is  $p=14$  and  $p=3$  for the cases (a) and (b).

stochastic processes at different temporal scales, and to a reliable estimation of such flow starting from simple AR identification. The proposed framework is flexible enough to encompass more general model representations that may unveil important multiscale features of coupled processes. For instance, integrating the standard AR representation with fractional integrated (FI) innovation modeling [37] would be straightforward as ARFI models have an SS representation, and would easily lead to assess multiscale GC in the presence of long-range correlations. The proposed setting provides also the basis to expand the applicability of multiscale GC to nonstationary and nonlinear SS processes [38], and to formalize exact computation of cross-scale information transfer within and between multivariate processes [17], thus opening new avenues of research in the evaluation of causal interactions among coupled processes.

[1] N. Wiener, *Modern mathematics for engineers* **1**, 125 (1956).  
[2] C. W. Granger, *Econometrica: Journal of the Econometric Society*, 424 (1969).  
[3] J. Geweke, *Journal of the American statistical association* **77**, 304 (1982).  
[4] J. F. Geweke, *J. Am. Stat. Ass.* **79**, 907 (1984).  
[5] J. R. Freeman, *American Journal of Political Science*, 327 (1983).  
[6] D. A. Smirnov and I. I. Mokhov, *Physical Review E* **80**, 016208 (2009).  
[7] S. L. Bressler and A. K. Seth, *Neuroimage* **58**, 323 (2011).  
[8] A. Porta and L. Faes, *Proceedings of the IEEE* (2015), 10.1109/JPROC.2015.2476824.  
[9] L. Barnett, A. B. Barrett, and A. K. Seth, *Phys. Rev. Lett.* **103**, 238701 (2009).  
[10] P. Ivanov, L. Nunes Amaral, A. Goldberger, S. Havlin, M. Rosenblum, Z. Struzik, and H. Stanley, *Nature* **399**, 461 (1999).  
[11] X. Kang, X. Jia, R. Geocadin, and N. V. Thakor, *IEEE Trans. Biomed. Eng.* **5**, 1023 (2009).  
[12] J. Valencia, A. Porta, M. Vallverdú, F. Clariá, R. Bara-

nowski, E. Orłowska-Baranowska, and P. Caminal, *IEEE Trans. Biomed. Eng.* **56**, 2202 (2009).  
[13] C.-M. Chou, *Entropy* **13**, 241 (2011).  
[14] J. Wang, P. Shang, X. Zhao, and J. Xia, *International Journal of Modern Physics C* **24**, 1350006 (2013).  
[15] M. Costa, A. L. Goldberger, and C.-K. Peng, *Phys. Rev. Lett.* **89**, 068102 (2002).  
[16] M. Lungarella, A. Pitti, and Y. Kuniyoshi, *Phys. Rev. E* **76** (2007).  
[17] M. Paluš, *Entropy* **16**, 5263 (2014).  
[18] L. Barnett and A. K. Seth, *Phys. Rev. E* **91**, 040101 (2015).  
[19] V. Solo, *Neural computation* **28**, 914 (2016).  
[20] L. Barnett and A. Seth, *J. Neurosci. Methods* **201**, 404 (2011).  
[21] M. Aoki and A. Havenner, *Econ. Rev.* **10**, 1 (1991).  
[22] B. D. Anderson and J. B. Moore, *Englewood Cliffs* **21**, 22 (1979).  
[23] Bad lowpass filtering such as the averaging may induce the detection of spurious influences; this was observed in Fig. 4 of our preliminary work L. Faes, A. Montalto, S. Stramaglia, G. Nollo, and D. Marinazzo, *arXiv preprint arXiv:1602.06155* (2016).  
[24] The developed filter is a FIR filter of order  $q$  designed with the classic window method applying the Hamming window; see e.g. A. V. Oppenheim and R. W. Schaffer, *Digital signal processing* (Prentice-Hall Englewood Cliffs, NJ, 1975).  
[25] in this study AR models were identified through ordinary least squares and using the Bayesian Information Criterion to set the model order as seen e.g. in S. L. Marple, *Digital spectral analysis: with applications*, Vol. 5 (Prentice-Hall Englewood Cliffs, NJ, 1987).  
[26] B. B. Booth, C. D. Jones, M. Collins, I. J. Totterdell, P. M. Cox, S. Sitch, C. Huntingford, R. A. Betts, G. R. Harris, and J. Lloyd, *Environmental Research Letters* **7**, 024002 (2012).  
[27] T. L. Delworth, A. J. Broccoli, A. Rosati, R. J. Stouffer, V. Balaji, J. A. Beesley, W. F. Cooke, K. W. Dixon, J. Dunne, K. A. Dunne, *et al.*, *Journal of Climate* **19**, 643 (2006).  
[28] E. Kodra, S. Chatterjee, and A. R. Ganguly, *Theoretical and Applied Climatology* **104**, 325 (2011).  
[29] A. Attanasio, *Theoretical and Applied Climatology* **110**, 281 (2012).  
[30] A. Stips, D. Macias, C. Coughlan, E. Garcia-Goriz, and X. S. Liang, *Scientific Reports* **6**, 21691 (2016).  
[31] “Nasa - goddard institute for space studies,” <https://data.giss.nasa.gov/>, accessed: 2017-03-23.  
[32] “Earth system research laboratory,” <https://www.esrl.noaa.gov/>, accessed: 2017-03-23.  
[33] “National oceanic and atmospheric administration,” <https://www.ncdc.noaa.gov/data-access/paleoclimatology-data/datasets/ice-core>, accessed: 2017-03-23.  
[34] J. Kang and R. Larsson, *Theoretical and Applied Climatology* **116**, 537 (2014).  
[35] N. Caillon, *Science* **299**, 1728 (2003).  
[36] M. Scheffer, V. Brovkin, and P. M. Cox, *Geophysical Research Letters* **33**, L10702 (2006).  
[37] R. J. Sela and C. M. Hurvich, *Journal of Time Series Analysis* **30**, 631 (2009).  
[38] G. Kitagawa, *Journal of the American statistical association* **82**, 1032 (1987).